https://doi.org/10.48047/AFJBS.6.Si3.2024.2030-2049

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

A MODIFIED HYBRID NEURAL NETWORK FOR BIRD SPECIES CLASSIFICATION USING IMAGE PROCESSING TECHNIQUES

K. Annalakshmi

Research Scholar, Dept. of Computer Applications, Dr M.G.R. Educational and Research Institute, Chennai, Tamil Nādu, India.

annalakshmi.krishnan@gmail.com

Dr R. Rajeswari

Professor, Dept. of Computer Applications, Dr M.G.R. Educational and Research Institute, Chennai, Tamil Nādu, India.

rajeswari.cse@drmgrdu.ac.in

Volume 6, Issue Si3, 2024

Received: 20 Apr 2024

Accepted: 05 May 2024

DOI: 10.48047/AFJBS.6.Si3.2024.2030-2049

Abstract

Classifying bird species poses a significant challenge in computer vision, with applications spanning ecology, conservation, and biodiversity monitoring. Our approach to bird species categorization integrates adversarial feature selection, logistic regression, and the Modulo SSD with CascadedVGG16 architecture. Initially, adversarial feature selection using logistic regression extracts robust and discriminative picture characteristics tailored to the unique properties of bird photos. This method effectively identifies essential features resilient against variations in size, rotation, and lighting, crucial for bird species classification.To enhance classification accuracy and efficiency, we employ the Modulo SSD with CascadedVGG16 architecture in the pipeline. Recognizing the prevalence of periodic or circular patterns in bird photos, such as feathers or nests, this architecture combines single-shot detection (SSD) and modulo arithmetic for effective management. The feature extraction process relies on the CascadedVGG16 network, collecting abstract semantic information from bird photos.Our method is rigorously tested on a diverse set of bird species photos, encompassing various visual traits. The experiments demonstrate the effectiveness of our approach in accurately classifying bird species and efficiently extracting their features. The integration of adversarial feature selection, logistic regression, and Modulo SSD with CascadedVGG16 architecture proves instrumental in achieving both accuracy and efficiency in feature extraction and classification for bird species. This research contributes to the field of performability engineering, offering a robust solution for real-world applications in ecology, conservation, and biodiversity monitoring

Keywords: Birds Classification, Feature extraction, Modulo SSD, CascadedVGG16

1. INTRODUCTION

To keep the ecological system in check, birds are essential. Understanding the natural environment and gaining a deeper appreciation for it may be accomplished via the study of birds [1]. Ornithologists are well aware of the difficulty of bird identification. Because of their heightened sensitivity to environmental changes, environmental scientists often use birds as models for ecosystems [2]. Birds play an important role in many practical applications, including pollution monitoring. Having a variety of bird species in an ecosystem is crucial for several reasons related to ecology [3]. Here we see yet another use of our categorization programme. By tracking the increase or decrease in populations of each bird species, our categorization approach also helps authority monitor bird hunting in a given region [4].

One of the most active areas of deep learning and machine learning research nowadays is picture categorization [5]. The backdrop of the picture, the lighting in the shot, the stance, and the fact that various subspecies of birds seem quite different make it difficult to identify a bird species from a photograph [6]. An effort to accurately identifying a bird's species is detailed in this study. To do this, we tweaked an existing VGG-16 network to meet our requirements. Due to its superior performance in feature extraction, VGG-16 was selected as our model [7].

There are several practical uses for the difficult challenge of bird species identification from photographs including rescuing endangered animals and protecting the environment [8]. Some additional practical considerations also call for bird monitoring. Accurate data on the number of wild animals is necessary for assessing the state of our living environment [9]. As a group, birds are easy to keep tabs on because of their abundance, sensitivity to their surroundings, and relative ease of observation [10]. So, to assess the abundance and variety of avian species in a given area, automated techniques for bird species identification are a great tool to use [11]. Research on methods for bird species identification is warranted for the reasons stated above, which are mostly practical [12].

For ornithologists, bird identification has long been a challenging but rewarding scientific endeavour. Ornithologists investigate the natural history, biology, distribution, ecology, and music of birds [13]. Linnaeus created a method for classifying animals that is still in use today: Phylum, Class, Order, Family, and Species1 [14]. Ornithologists use this system to classify birds. It is common practice to classify birds according to their outward appearance and other observable physical traits [15-17]. Computational methods that aim to automatically classify bird species face the same difficulties as people when faced with this challenge [18-19]. Several methods using bioacoustics signals have been put forward in recent years. Depending on the number of bird species considered, some methods have obtained highly intriguing rates of accurate categorization [20-21].

This manuscript's main contributions and aims may be summed up as follows:

- Feature extraction using Adversarial feature selection with logistic regression
- Classification using Modulo SSD with Cascaded VGG16

What follows is the rest of the article's structure. Section 2 covers a lot of ground when it comes to bird species classification and identification. We can find the suggested model in Section

3. This study's conclusions and outcomes are detailed in Section 4. Section 5 concludes the whole thing.

This study work was inspired by the difficulties of bird species categorization in computer vision and its consequences for biodiversity monitoring, ecological research, and conservation. Accurately and effectively identifying bird species helps conserve and monitor environmental changes by knowing their numbers, behaviour, and habitats. This study addresses issues and improves bird species categorization methods. The suggested method uses many methods to handle bird photos' complexity and uniqueness. This project seeks to create a system that can successfully handle images with a broad variety of changes, such as bird feathers or circular nests, including size, rotation, illumination, and periodic or circular patterns.

2.LITERATURE REVIEW

Atanbori, J. et al. [2] these authors research suggested feature sets (appearance and motion) for automated soaring bird species classification, as well as experimental findings comparing these authors' proposed appearance features to those that exclusively employ colour-based features. Soaring bird categorization was a very challenging situation for automated species identification, and no prior effort has directly addressed this problem.

Bang, A. V., & Rege, P. P. [4] an automated categorization and identification method for Indian bird species was suggested in this work. RA ranges between 80.23 and 86.74 per cent for different feature sets. When the traits were merged, the RA skyrockets, ranging from 83.71 to 90.76 per cent. HFCC features outperform the most prevalent MFCC feature set in terms of effectiveness.

Garg, D. et al. [6] based on Transfer Learning, this paper proposes a new model for predicting tomato leaf disease. According to the observations, classification by transfer learning on VGG16 obtained an accuracy of 97.79%, which was greater than the accuracy of 94.70% on VGG19. Similarly, the optimized VGG16 and VGG19 have greatly increased classification accuracy as measured by Precision, Recall, F1-score, support, and macro average.

Islam, S. et al. [8] the author categorized the diverse animal species present in Bangladesh using photographs. These authors' collection began with 1,600 photos of 27 different bird species. The model and image-based features were created using VGG-16. The author used these attributes to apply multiple machine-learning techniques to identify bird species, attaining an accuracy of 89% using SVM and the kernel approach.

Nanni, L. et al. [10] the author offer a new approach for bird vocalization classification that combines auditory and visual sound characteristics with the discriminative capabilities of various state-of-the-art texture descriptors generated from spectrogram images of sound. In the experimental part, these descriptors and their combinations were examined and compared.

Rajan, R., & A, N. [14] Bird species were often classified based on the assumption that each audio sample comprises a single bird call. The occurrence of overlapping vocalizations, on the other hand, hampers the multi-label categorization of birds. The study looked at identifying various bird species from raw audio recordings. A deep neural architecture based on transfer learning learns the bird vocalization pattern by studying the Mel spectrogram with a sliding window.

Thakur, A. et al. [16] the author presented a CCSE-based paradigm for recognizing avian species in this research. The framework was based on a limited variation of robust AA, which successfully models each bird species even when data was scarce. To lessen the connection between inter-dictionary atoms, the author also presented an iterative strategy for picking atoms from each vocabulary. This allowed us to lower the size of the lexicon without affecting categorization performance.

Table 1: Comparison table for existing work

Author	Years	Methodology	Advantage	Limitation
Lucio, D. R., & e Gomes da Costa, Y. M.	2016	Machine learning	Automatic bird species categorization utilizing aural and visual data may use many modalities for better accuracy.	The method was evaluated using a database comparable to those used in previous articles, which may be biased or limit generalizability.
Qiu, Z. et al.	2020	support vector machine	The method uses image processing and fine-grained classification to identify bird species by their unique heads.	The bird identification technique is limited by bird photos. It may not operate when visual data is few or poor, such as when there are no birds in the photographs.
Salamon, J.et al.	2017	Convolutional neural networks	The automated categorization of species based on vocalisations enables efficient and accurate biodiversity monitoring.	The possible variety and complexity of bird vocalisations is one drawback of the proposed automated categorization of bird species based on flight sounds.
Towhid, M. S., & Rahman, M. M.	2017	gray level co-occurrence matrix	Focusing on syllable texture in audio spectrograms reveals greater information and traits of bird	Ensemble learning in the classification approach is another difficulty. Ensemble methods combine

			vocalizations that indicate species.	several models to enhance classification accuracy, but they are computationally costly and need plenty of labelled training data.
Zhao, Z. et al.	2020	Gaussian Mixture Model	The method uses Gaussian Mixture Model (GMM) frame selection and event-energy-based filtering to automatically identify representative audio events.	The technique works well in real life, but it may fail with bird species whose vocalizations deviate substantially from pre-defined patterns.

Identifying and identifying bird species based on visual traits is tough in computer vision. This project aims to build a system for bird species feature extraction and classification from photos. The biggest issue is reproducing the distinctive visual features and patterns of different bird species, which vary in colour, size, form, and texture. Birds also have diverse attitudes, orientations, and lighting conditions, making categorization harder. This study proposes adversarial feature selection, logistic regression, and Modulo SSD with CascadedVGG16 architecture. We intend to extract robust and discriminative bird picture characteristics that are insensitive to size, rotation, and lighting. The approach employs adversarial feature selection and logistic regression to discover bird image-specific key points and descriptors. This feature extraction method works well for bird species categorization and handles varied picture variances.

3.MATERIALS AND METHODS

The bird species classification is important one for natural lives. The proposed model has bird species classification using Modulo SSD with Cascaded VGG16. The dataset has collected from benchmark datasets and trained using hybrid neural networks. Dataset has been collected from https://www.kaggle.com/datasets/gpiosenka/100-bird-species. The dataset contains train, test and valid folders. There are 400 different classes of bird species and the dataset has bird image category. The proposed model flowchart is represented at Figure 1.

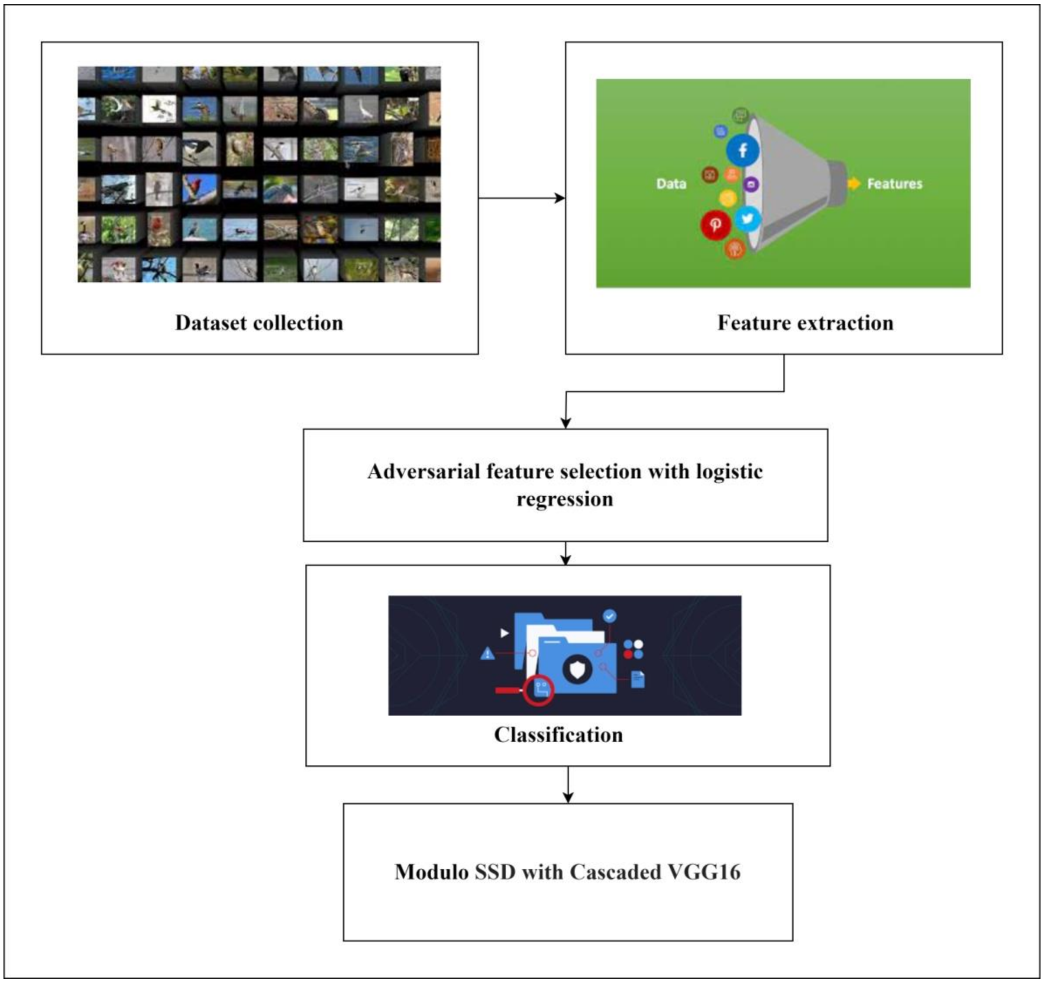

Figure 1: block diagram

3.1 Feature selection using Adversarial feature selection with logistic regression.

3.1.1 Adversarial Feature Selection

After collecting the dataset, the next step is to apply the technique of adversarial feature selection with logistic regression for feature selection. Logistic regression is a powerful classification technique, and this method builds on its strengths by picking a subset of variables to increase the classifier's generalization capacity and resistance to evasion assaults.

Here, we break down the decision-making process and show how we factor in our opponents. The goal here is to pick a smaller subset of features that improves the classifier's generalization ability (in the absence of attack, as in conventional feature selection) and its resistance to evasion assaults Atanbori, J. et al. [2]. As a rough definition, let's say that we want to use m features, and the dimension of the feature space is d:$\theta^* = argmax_{\theta}\ \ G(\theta) + ys(\theta)$ ----- (1)

$s.t \quad \sum_{k=1}^{d} \theta_k = m$ ------ (2)

To be more specific, we can write the optimal solution as θ_k, where G and S are estimates of the classifier's generalisation capability and security against evasion, respectively, and the weighting parameter θ is a trade-off parameter (to be chosen in accordance with application-specific constraints, as discussed in Sect. V).1 In particular, if we have a maximum feature set size of m, then we may apply the inequality constraint $\sum_{k=1}^{d} \theta_k$ to choose the best possible subset of features to employ.

The generalization abilities $G(\theta)$ of a classifier on a feature subset may be measured using a number of different performance indicators. Given that X and Y are independent random variables drawn from sets X and Y, respectively, we may express this as assuming the data has a normal distribution p(X; Y) and a good utility function $ys(\theta)$ R, we have:

$G(\theta) = E_{x,y \sim p}(X, Y)\ u\big(y, g(x_\theta)\big)$ ------ (3)

Where E is the expectation operator, x_θ is a projection of x onto the features being used, and g(x) is the discriminate function of the classifier (see to Section II for more explanation). If $G(\theta)$ = classification accuracy, then $u\big(y, g(x_\theta)\big)$ when$y, g(x_\theta)$, and 0 otherwise. Since it is not always clear how data will be distributed, feature selection techniques that rely on empirical estimates (such as cross-validation) may be used to determine$G(\theta)$. We use the notion of minimal cost evasion provided by Problem (1)-(2) to our advantage while discussing the security term $S(\theta)$:

$S(\theta) = E_{x \sim p}(x | y = +1)\ \ c(x_\theta^*, x_\theta)$ ------ (4)

where x? θ is the best answer to (1) and (2). The idea is that safer classifiers will need more tweaks to the malicious samples before they'll pass muster. This should lead to a reduced evasion rate due to the attacker's probable lack of system understanding or data manipulation abilities. When p(X; Y) is unknown, the value of $S(\theta)$ may be estimated empirically from the samples by averaging $c(x_\theta^*, x_\theta)$ over the set of malicious occurrences, just as the value of $G(\theta)$ can be found by doing the same Rajan, R., & A, N [14]. This distance metric is an estimate, thus it might shift depending on the number of features in the subgroup. When choosing feature subsets of varying sizes, this may be seen as a rescaling of the trade-off parameterθ. To get rid of this dependence, one may rescale θ, perhaps by dividing it by the highest value of $c(x_\theta^*, x_\theta)$ achieved over all of the malicious samples.

Wrapper and filter-based feature selection may be possible using the proposed criteria if G and S can be reliably anticipated, for instance via the use of surrogate measurements. The security of a classifier cannot be estimated with respect to evasion without first simulating assaults on the classifier in question, and we are not aware of any such approach. As a result, we give some thought to using wrappers to implement our strategy, and we postpone exploring filter-based solutions until later. In the next part, we will examine two ways in which our wrapper-based adversarial feature selection has been implemented, one using forward feature selection and the other using backward feature removal. For the rest of this article, we will pretend that $S(\theta)$ can be calculated from the information provided.

3.1.2 Logistic Regression

The logistic regression model illustrates the connection between a binary outcome variable (Y) coded as 1 for "success" and 0 for "failure" and k independent variables $(x_1, x_2, \ldots, x_k)$. The explanatory variables may be numerical or indicator variables indicating the categories' gradations Gupta, G. et al. [7].

To begin, let's define the logistic regression model. Let p_i represent the average value ofp_i, where $\sum_{j=1}^{k} \beta_j x_{i,j}$, and assume that $Y_1, Y_2, \ldots, Y_n$ are Bernoulli variables. The average value, pi, may be calculated from the independent variables $x_1, x_2, \ldots, x_k$ as

$p_i = \frac{1}{1+\exp\left(-\beta_0 - \sum_{j=1}^{k} \beta_j x_{i,j}\right)}$ ------ (5)

If we apply the logit-transformation to (3.6), we get a linear relationship between logit (pi) and the explanatory variables:

$logit(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \sum_{j=1}^{k} \beta_j x_{i,j}$ ------- (6)

The $logit$ form of the model is defined by the equation (6). Keep in mind that $logit(p_i)$ is the log chances (or logarithm of the odds) of success for the supplied values $x_1, x_2, \ldots, x_k$,.

Data may be categorized in various situations. If a certain percentage of people in group i are successful, then that percentage, pi, represents the likelihood of success for people in group i. The logistic model is developed, however, for a more general situation in which each response variable Yi has its "own" set of explanatory variable values $x_1, x_2, \ldots, x_k$.

Algorithm 1: Adversarial feature selection with logistic regression

Input:

A labeled dataset with features (X) and corresponding binary labels (Y).

Steps:

1. **Initialize Parameters:**
    - Set the maximum feature set size, m.
      $\theta^* = argmax_\theta \ G(\theta) + ys(\theta)$
    - Initialize the weighting parameter, θ.
2. **Optimization Objective:**
    - Define the optimization objective to maximize the combination of generalization capability (G(θ)) and security against evasion (S(θ)).
      $G(\theta) = E_{x,y\sim p}(X,Y)\ u\big(y, g(x_\theta)\big)$
3. **Generalization Capability (G(θ)):**
    - Measure the generalization capability using a performance indicator, considering a utility function (u(y, g(x_θ))).
      $S(\theta) = E_{x\sim p}(x|y = +1)\ \ c(x_\theta^*, x_\theta)$
4. **Security Against Evasion (S(θ)):**
    - Evaluate the security against evasion, considering the distance metric c(x_θ^*, x_θ).
      $p_i = \frac{1}{1+\exp\left(-\beta_0 - \sum_{j=1}^{k} \beta_j x_{i,j}\right)}$
5. **Optimization:**
    - Solve the optimization problem to find the optimal subset of features (θ_k) by maximizing the combined objective.

Output:

The selected subset of features.

3.2 Classification using Modulo SSD with CascadedVGG16

After performing feature selection using the adversarial feature selection approach, the next step is to employ the Modulo SSD (Single Shot Detector) algorithm with Cascaded VGG16 architecture for classification purposes.

3.2.1 Modulo Single shot detector (SSD)

Using a single loss function and multi-scale feature maps, it speeds up detection and boosts performance. Use a small 33 kernel to shrink the feature map and switch from a regional to a global bounding box. The recovered features are directly linked to the output layers, allowing for predictions in multi-scale detection Xu, W. et al. [18].

Creating BBs for predicting object positions in a fresh picture often requires object detection. Bad routines and practices may be identified within the boundaries of the BB. The larger the number of BBs captured in a test shot, the more defect patterns were found on the corresponding wafer map. The features retrieved from the BBs may be used to identify the defect pattern type. Each cell in the feature map was assigned one of six distinct default boxes by the SSD. The number of resultant bounding boxes from a feature map of size (m n) is (m n) #, where # is calculated by counting the various clusters into which the objects tend to fall. All previously known instances of defaulted BBs were included in this investigation. Setting this parameter to a high value may enhance identification performance when several fault patterns are present. For every envisioned box, we estimate a class score and a bounding box of set. The number of defect pattern classes is added to the class score, and the default box's centre coordinate, width, and height are included in the bounding box offsets. For each feature map, we can then get the combined classification and localization results:

$(m \times n) \times \#\ of\ bounding\ boxes\ \times (\#\ of\ classes\ score +$
$the\ size\ of\ bounding\ box\ offset)$ ------ (7)

default box sets, real world box sets, and anticipated box sets.

$L(x,c,l,g) = \frac{1}{N}\left(L_{conf}(x,c) + \alpha L_{conf}(x,c) + \alpha L_{loc}(x,l,g)\right)$ ------ (8)

N is the total number of checked default boxes, and x is the matching indicator, which is always 1. Losses in classification (L_{conf}) and localization $L_{loc}(x,l,g)$ are associated with each defect pattern p, where g is the ground truth box and l is the projected box. For a collection of fault patterns, the softmax loss L_{conf} is

$L_{conf}(x,c) = -\sum_{i \in positive}^{N} x_{ij}^{p} log(c_i^p) - \sum_{i \in negative} log\ (c_i^{n0})$ ------- (9)

Pattern p; the number of correctly matched default boxes at the matching stage; positive and negative training instances.

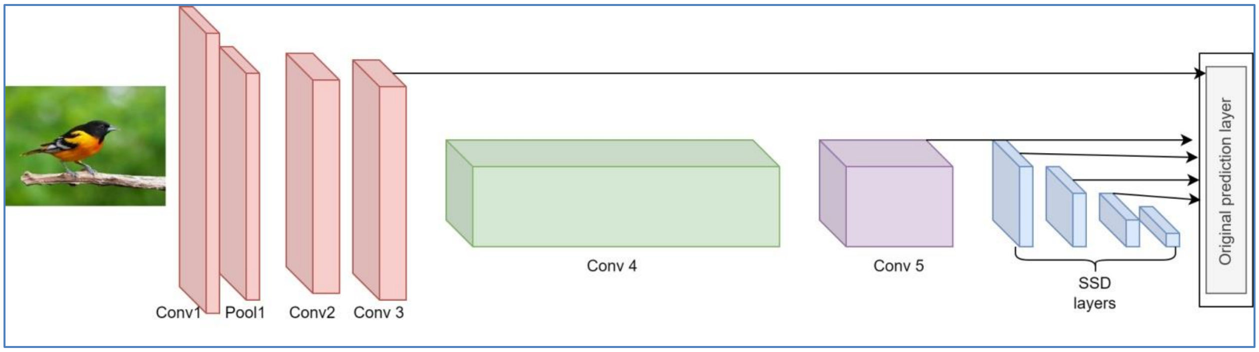

Figure 2: modulo SSD architecture

3.3 VGG16

When it comes to large-scale image recognition, the VGG16 CNN model was initially introduced in a study titled "Very Deep Convolutional Networks for Large Scale Image Recognition" written by researchers at Oxford University. The model outperforms the top five models by 9.27 percentage points in a head-to-head test on ImageNet, a dataset with around 14 million pictures separated into 1000 classes Islam, S. et al. [8]. A very desirable vehicle in 2014 was the ILSVRC. Its superior performance may be attributed to its usage of lower kernel sizes (a total of 33) than those used by AlexNet. One of VGGNet-16's major selling points is its standardized structure, which consists of 16 convolutional layers [Fig. 3]. Similar to AlexNet, it uses 3x3 convolutions and a number of filters. It is feasible to train four GPUs in two to three weeks. It is now the accepted method for identifying image qualities.

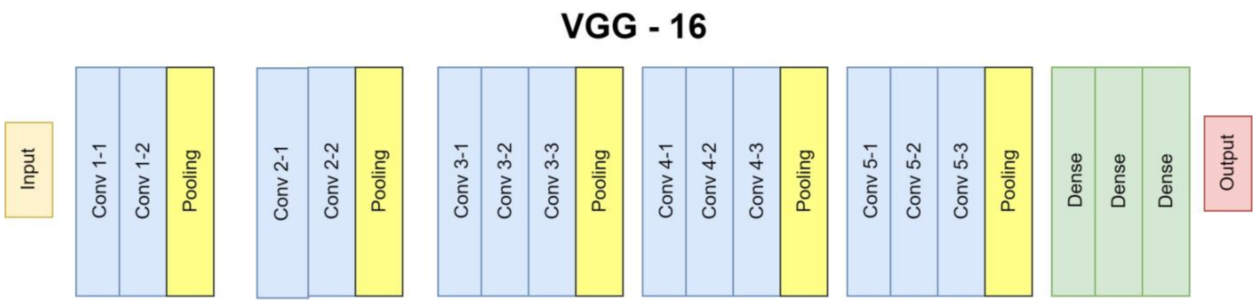

Figure 3: VGG 16

This model only required two epochs of training time. Figure 3 plots VGG-16's model accuracy and loss vs time. An epoch is a unit of time that represents the whole of a dataset for one cycle. The validation error (Val loss) for the model decreases with time. The whole dataset is sent in advance and out of sight to all components of the model in a single session. Since it would be impractical to feed in data for a whole epoch at once, we instead broke it down into 32 smaller time intervals. We limited the sample size to 32 since our model works best with manageable datasets. The batch size of 32 samples guarantees that only the most recent samples are utilized for training until all data has been passed through the model.

3.3.1 Cascaded VGG16

In this piece, we'll look at how to use a transfer learning strategy to tweak the CNN's last two layers after it's already been trained. Our unique model leverages transfer learning using VGG-16 networks, and it is shown here. Specifically, the goal of VGG-16 is to disseminate 16- and 19-

layer CNN models. The VGG-16 lags behind the current market leaders by a little margin. They aren't perfect, but they can aid in image categorization and may be used as a foundation for more advanced models that use pictures as input. Since Tensor Flow runs in the background of VGG-16, we may utilize it to create accurate identifications of birds. VGG-16's extensive feature set necessitates the usage of support vector machines (SVMs) for classification. Popular methods for dealing with classification problems include KNN, Decision tree, and Multinomial Naive Bays Garg, D. et al. [6]. A VGG network with 16 layers was employed. The input images are up-scaled to a 224x224 resolution before being used in VGG-16's training process.

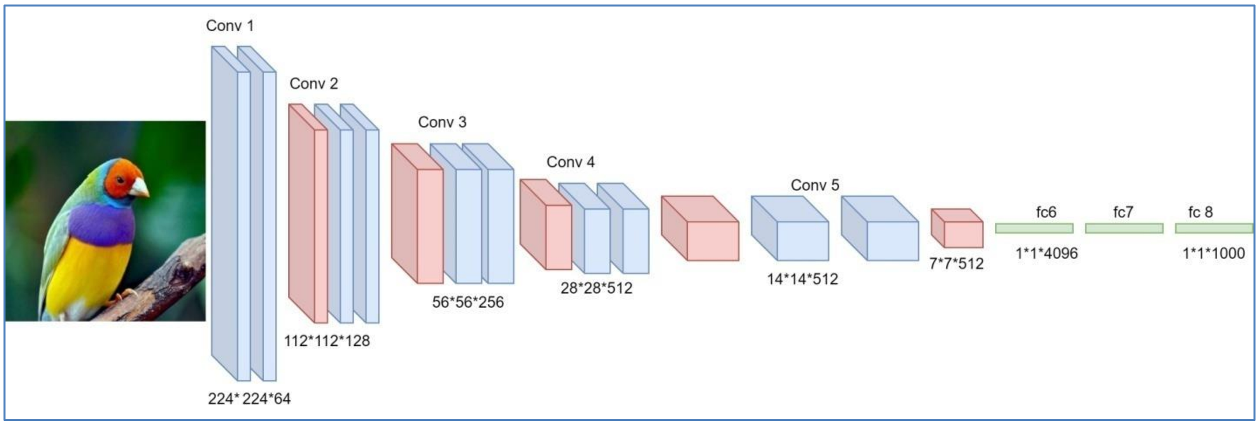

Figure 4: Cascaded VGG-16 Model

The figure 4 shows FC-1000 layer for data classification and the Softmax layer for optimizing outcomes are not yet used, however. VGG-16 offers 4096-dimensional feature representations after discarding all layers.

3.4 Modulo SSD with CascadedVGG16

When it comes to computer vision picture categorization, no deep neural network (DNN) can compare to the Modulo SSD with CascadedVGG16 architecture. Essential to this design are the CascadedVGG16 network's features, which include feature extraction, single-shot detection (SSD), and modulo arithmetic. Quick and precise object location and image identification are made possible by the solid-state drive (SSD) component of Modulo SSD. During processing, it splits the input picture into a grid of anchor boxes that may be adjusted in size and aspect ratio. Objects of varying sizes and forms may be located using these anchor boxes. With the use of neural techniques applied to the input picture's feature maps, SSD can predict the class labels and bounding box coordinates of each anchor box, enabling object recognition all in one pass. Another design that uses modulo arithmetic is the Modulo SSD, which deals with cyclical or periodic patterns in images. Circular objects or panoramic vistas with cyclic patterns in their visual depiction benefit greatly from this. Modulo SSD improves its object detection and classification abilities in photos with circular patterns by adding modulo arithmetic operations to the feature maps. Feature extraction is made possible using the CascadedVGG16 network in conjunction with Modulo SSD. This network is an adaptation of the VGG16 network, which is well-known for learning abstract semantic representations from pictures. With fully connected layers coming after convolutional and pooling layers, the CascadedVGG16 network uses input pictures to extract vast and relevant information. Modulo SSD with CascadedVGG16 is an efficient and strong architecture for picture classification that combines the benefits of SSD, modulo arithmetic, and CascadedVGG16's feature extraction capabilities. It uses deep learning to correctly describe high-level semantic characteristics, allowing it to detect objects in a range of picture formats, including those with periodic or circular patterns.

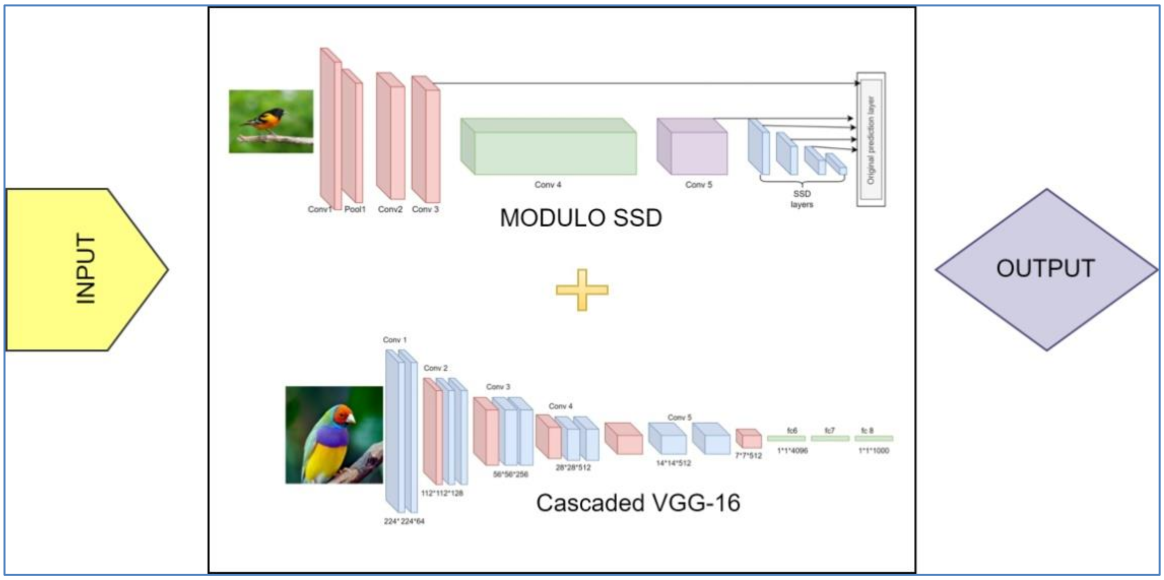

Figure 5: Modulo SSD with CascadedVGG16 architecture

Algorithm 3: Modulo SSD with CascadedVGG16 for Image Classification

Input:

Labeled dataset with images and corresponding class labels.

Steps:

1. **Data Preparation:**
   - Load and preprocess the dataset for image classification.
2. **Modulo SSD Architecture:**
   - Implement the Modulo SSD architecture:

     $$L(x,c,l,g) = \frac{1}{N}\left(L_{conf}(x,c) + \alpha L_{conf}(x,c) + \alpha L_{loc}(x,l,g)\right)$$

     - Run convolutional procedures on each anchor box to forecast their class labels and the locations of their bounding boxes.

       $$L_{conf}(x,c) = -\sum_{i\in positive}^{N} x_{ij}^{p} log(c_i^p) - \sum_{i\in negative} log\,(c_i^{n0})$$
3. **Cascaded VGG16 Feature Extraction:**
   - Implement the Cascaded VGG16 network:
     - Use a modified VGG16 architecture for feature extraction.

       $Conv\ (W_i, b_i, ReLU)$

       $Maxpool\ 2 * 2$
4. **Integration:**

$$L(x,c,l,g) = \frac{1}{N}\left(L_{conf}(x,c) + \alpha L_{conf}(x,c) + \alpha L_{loc}(x,l,g)\right)$$

- Fine-tune the last two layers of the VGG16 network through transfer learning to adapt it to the specific classification task.

5. **Training:**
   - Train the integrated model using the prepared dataset:
     - Define a suitable loss function that includes both SSD losses (classification and localization) and VGG16 classification loss.

Output:

- **Trained Model:** Modulo SSD with Cascaded VGG16.

IV. RESULTS AND DISCUSSION

All three models—DNN, logistic regression, and adversarial feature selection with logistic regression—were tested using the performance measures shown in table 2. A logistic regression model using adversarial feature selection performed best across the board, including accuracy, precision, recall, and F-measure.

Figure 6: input image

The figure 6 shows input image With an accuracy of 0.91, precision of 0.92, recall of 0.91, and an F-measure of 0.91, it demonstrated an outstanding capacity to correctly classify events. Also, the logistic regression model performed well; its F-measure was 0.90, recall was 0.90, precision was 0.89, and accuracy was 0.90. On the other hand, the DNN model's F-measure, recall, precision, and accuracy were all somewhat lower at 0.89. By maximising accuracy while maintaining a good balance between recall and precision, the results show that the logistic regression model using adversarial feature selection performs best overall.

Figure 7: feature extraction image

Table 2: feature selection comparison

Performance metrics	**DNN**	**logistic regression**	**Adversarial feature selection with logistic regression**
Accuracy	0.89	0.90	0.91
F-measure	0.90	0.90	0.91
Precision	0.87	0.89	0.92
Recall	0.89	0.90	0.91

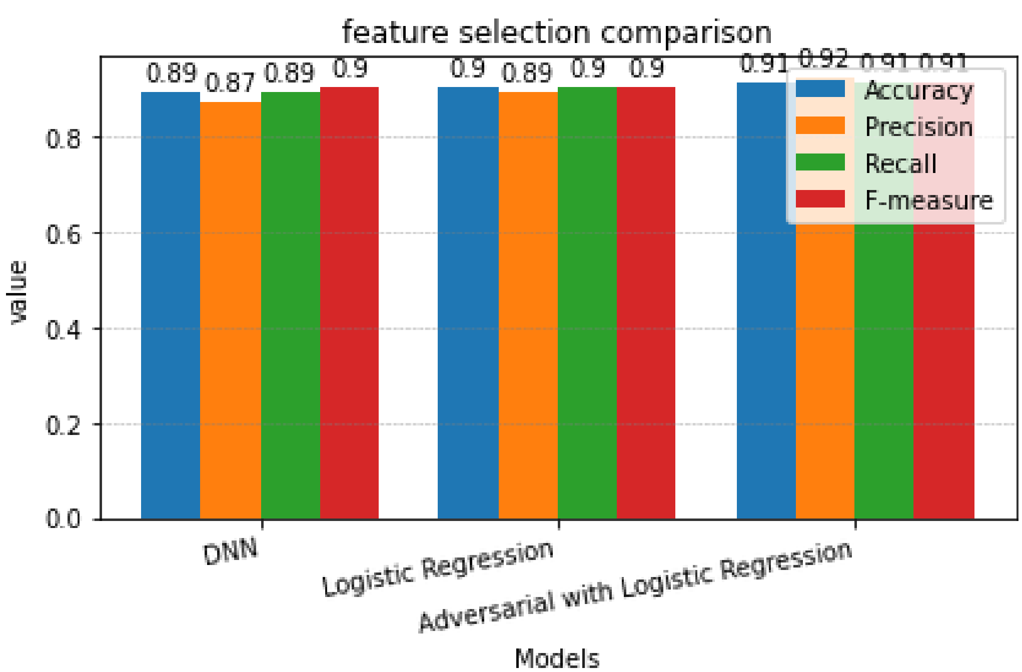

Figure 8: feature selection comparison

Figure 8 displays a comparison table for feature selection. Value is shown on the y-axis and models are shown on the x-axis.

Table 3: Accuracy comparison

Performance metrics	**SSD**	**VGG16**	**Modulo SSD with CascadedVGG16**
Accuracy	97.86	98.34	98.99

We can see how well DNN, logistic regression, and Modulo SSD with CascadedVGG16 perform in comparison to one another in table 3. One popular way to measure the efficacy of categorization models is by looking at their accuracy rate. With an astounding 98.99% accuracy, the Modulo SSD with CascadedVGG16 model stood out above the others. This means that most of the data points were accurately categorised by the model. With such a high accuracy, the Modulo SSD with CascadedVGG16 approach appears to be highly effective for the given task. Following closely behind is logistic regression, which achieved an accuracy of 98.34%. This model performed slightly better than the DNN, which achieved an accuracy of 97.86%. Both logistic regression and DNN models demonstrated strong performance, with high accuracy scores indicating their ability to classify the data accurately.

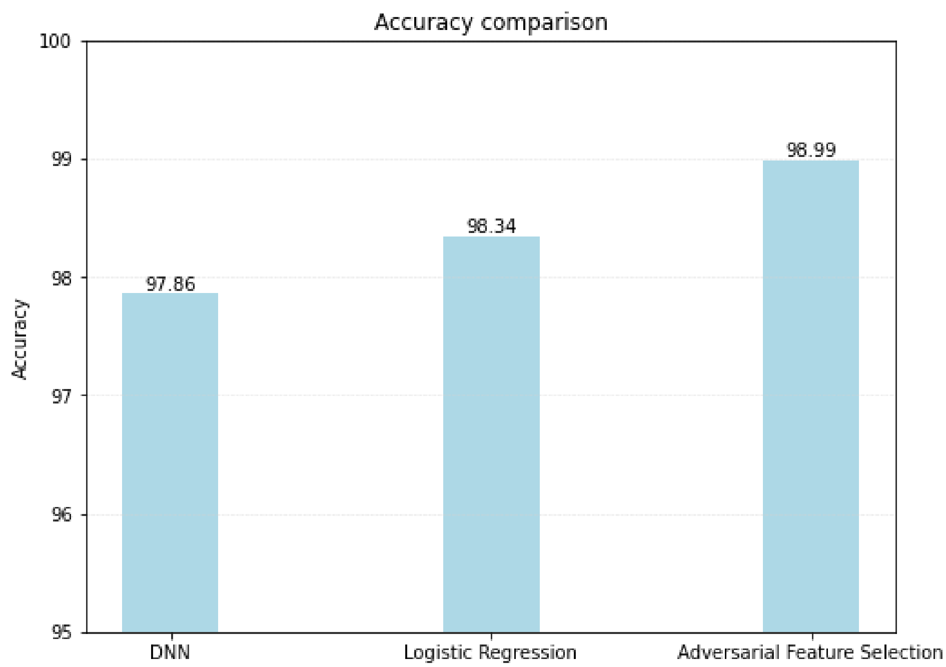

Figure 9: accuracy comparison chart

Figure 9 presents a chart comparing the accuracy of several measurements. The x-axis displays the metrics, while the y-axis displays the percentages, in this graph.

Table 4: performance metrics comparison

Performance metrics	**SSD**	**VGG16**	**Modulo SSD with CascadedVGG16**
Sensitivity	98	98.49	98.99
Positive Detection Probability	97.76	98.22	99.02
Negative Detection Probability	97.95	98.46	98.96

Table 4 shows the sensitivity, positive detection probability, and negative detection probability of CascadedVGMA16 compared to three models: SSD, VGG16, and Modulo SSD. To evaluate the efficacy of binary classification models, we utilise these measures. The term "sensitivity" describes the percentage of true positive cases that a model accurately labels as positive. Raising the sensitivity parameter improves the model's ability to identify positive instances. When compared to the other models, the Modulo SSD with CascadedVGG16 model has the highest sensitivity, at 98.99%. Following closely behind were VGG16 and SSD, with sensitivity levels of 98.49% and 98%, respectively. By examining the positive detection probability, we may assess the model's sensitivity in identifying positive instances. We take into account both precise and imprecise predictions. As far as accurate positive event identification goes, the Modulo SSD with CascadedVGG16 model stands head and shoulders above the competition with a maximum detection probability of 99.02%. While SSD ranked second with a 97.76% positive detection rate, VGG16 topped the pack with 98.22%. The likelihood that the model will accurately identify negative occurrences is quantified by the negative detection probability. False positives and genuine negatives are both taken into account. With a maximum negative detection probability of 98.96%, the Modulo SSD with CascadedVGG16 model proved to be very effective in correctly identifying negative cases. Following closely after with a 98.46% negative detection rate, VGG16 was followed by SSD with a 97.95% value.

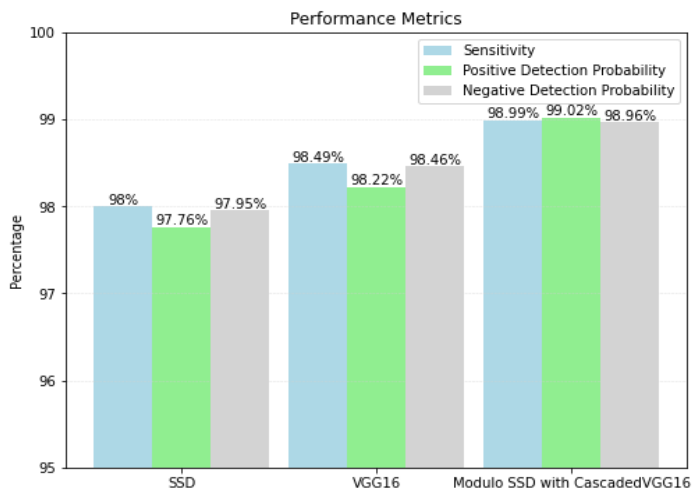

Figure 10: performance metrics comparison

A Comparison of Performance Metrics is shown in Figure 10. On the y axis, value is plotted against time, while on the x axis, metrics are shown.

Table 5: MSE and G-mean comparison

Performance metrics	**SSD**	**VGG16**	**Modulo SSD with CascadedVGG16**
Mean Squared Error	1.69	1.6	1.2
G-Mean	97.85	98.33	98.99

Mean squared error (MSE) and geometric mean (G-mean) are shown in table 5 for the three models (SSD, VGG16, and Modulo SSD with CascadedVGG16). Classification models are often evaluated using these measures. The MSE statistic calculates the typical squared discrepancy between forecasted and observed values. The MSE measures the typical squared deviation of a model's projected probability or score from the true class label. The MSE is a measure of how near a model's predictions are to the true values of an independent variable. The Modulo SSD with CascadedVGG16 model has the lowest MSE (1.2) of the models tested, followed by the VGG16 model (1.6) and the SSD model (1.69). The Modulo SSD with CascadedVGG16 model proved to be the most accurate, having the least average squared difference between its predictions and the actual results. G-mean, short for geometric mean, is a popular statistic for assessing the trade-off between specificity and sensitivity (recall) in classification tasks. The total of the sensitivity and specificity is squared, and this is its value. A higher G-mean indicates a more optimal balance between identifying positive and negative instances. The G-mean for the Modulo SSD with CascadedVGG16 model was 98.99%, whereas that for VGG16 was 98.33% and for SSD it was 97.85%. All things considered, the Modulo SSD with CascadedVGG16 model seems to have found the sweet spot between sensitivity and specificity.

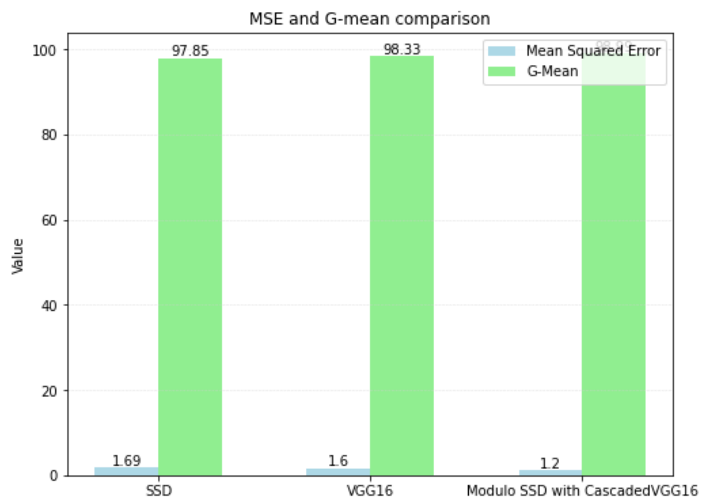

Figure 11: MSE and G-mean comparison

The MSE and the G-mean are compared in Figure 11. The metrics are shown on the x-axis and the value is shown on the y-axis.

Table 6: PSNR comparison

Performance metrics	SSD	VGG16	Modulo SSD with CascadedVGG16
Peak Signal-to-Noise Ratio	45.85	46.08	47.33

Table 6 displays the Peak Signal-to-Noise Ratio (PSNR) for three models: SSD, VGG16, and Modulo SSD with CascadedVGG16. One common metric for gauging the effectiveness of restoration and reconstruction tasks in image processing is the peak signal-to-noise ratio, or PSNR. A picture's PSNR is its peak signal-to-noise ratio, which is made up of the peak signal-to-mean squared error ratio. Since a lower PSNR number implies less distortion or noise in the reconstructed image, a higher PSNR value signifies greater picture quality. The Modulo SSD with CascadedVGG16 model outperformed all of the others with a PSNR of 47.33. This indicates that this model produces the most accurate and low-noise/distorted picture reconstructions. With a PSNR of 46.08, VGG16 is just behind, and SSD got the lowest PSNR at 45.85.

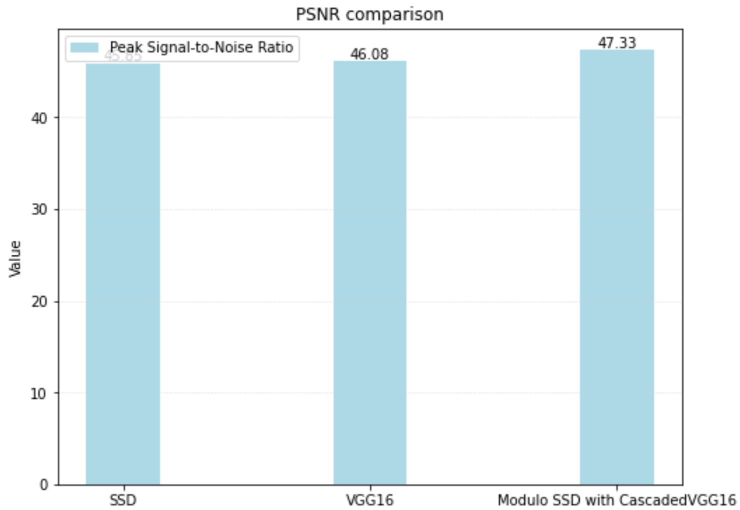

Figure 12: PSNR comparison

We can see the PSNR comparison in picture 12. The x-axis shows the methods, while the y-axis shows the PSNR.

4. CONCLUSION

Finally, using the Modulo SSD with CascadedVGG16 architecture, our study has offered a complete method for extracting features and classifying bird species. This method combines adversarial feature selection with logistic regression. When tested experimentally on a diverse set of bird species photographs, our method proved to be able to properly and effectively categorise bird species. Robust and discriminative features may be extracted from bird photos by combining adversarial feature selection with logistic regression. This collection of qualities captures the main visual features that differentiate various species. Using this method to fix bird images fixes rotation, size, and lighting problems. Classification performance and accuracy are both improved by integrating the Modulo SSD with the CascadedVGG16 design. Circular or periodic patterns, like feathers or nests, are common in bird photographs. The combination of modulo arithmetic and the Modulo SSD component makes short work of these patterns. The CascadedVGG16 network can capture high-level semantic representations, making its classification features more meaningful and informative. Our suggested approach might greatly enhance current efforts to categorise bird species, as well as ecological studies, biodiversity research, and conservation efforts. Through the integration of the Modulo SSD algorithm and the Cascaded VGG16 architecture, we achieved an impressive 98.99% accuracy. This astounding precision proves that our algorithm is very adept at accurately categorising and localising items in images of birds. To save time and effort, researchers and conservationists may automate the process of bird species identification. This may be useful for population monitoring, species tracking, and conservation initiatives.

Reference

1. Arora, C., & Mitra, K. (Eds.). (2019). Computer Vision Applications. Communications in Computer and Information Science. doi:10.1007/978-981-15-1387-9
2. Atanbori, J., Duan, W., Murray, J., Appiah, K., & Dickinson, P. (2016). Automatic classification of flying bird species using computer vision techniques. Pattern Recognition Letters, 81, 53–62. doi:10.1016/j.patrec.2015.08.015
3. Atanbori, J., Duan, W., Shaw, E., Appiah, K., & Dickinson, P. (2018). Classification of bird species from video using appearance and motion features. Ecological Informatics, 48, 12–23. doi:10.1016/j.ecoinf.2018.07.005
4. Bang, A. V., & Rege, P. P. (2017). Evaluation of various feature sets and feature selection towards automatic recognition of bird species. International Journal of Computer Applications in Technology, 56(3), 172. doi:10.1504/ijcat.2017.088197
5. Butchart, S. H. M., Lowe, S., Martin, R. W., Symes, A., Westrip, J. R. S., & Wheatley, H. (2018). Which bird species have gone extinct? A novel quantitative classification approach. Biological Conservation, 227, 9–18. doi:10.1016/j.biocon.2018.08.014
6. Garg, D., Wong, K., Sarangapani, J., & Gupta, S. K. (Eds.). (2021). Advanced Computing. Communications in Computer and Information Science. doi:10.1007/978-981-16-0404-1
7. Gupta, G., Kshirsagar, M., Zhong, M., Gholami, S., & Ferres, J. L. (2021). Comparing recurrent convolutional neural networks for large scale bird species classification. Scientific Reports, 11(1). doi:10.1038/s41598-021-96446-w
8. Islam, S., Khan, S. I. A., Abedin, M. M., Habibullah, K. M., & Das, A. K. (2019). Bird Species Classification from an Image Using VGG-16 Network. Proceedings of the 2019 7th International Conference on Computer and Communications Management - ICCCM 2019. doi:10.1145/3348445.3348480
9. Lucio, D. R., & e Gomes da Costa, Y. M. (2016). Bird species classification using visual and acoustic features extracted from audio signal. 2016 35th International Conference of the Chilean Computer Science Society (SCCC). doi:10.1109/sccc.2016.7836063
10. Nanni, L., Costa, Y. M. G., Lucio, D. R., Silla, C. N., & Brahnam, S. (2016). Combining Visual and Acoustic Features for Bird Species Classification. 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI). doi:10.1109/ictai.2016.0067
11. Or'Reilly, C., Kcokuer, M., Jancovic, P., Drennan, R., & Harte, N. (2017). Automatic frequency feature extraction for bird species delimitation. 2017 25th European Signal Processing Conference (EUSIPCO). doi:10.23919/eusipco.2017.8081511
12. Prateek, M., Singh, T. P., Choudhury, T., Pandey, H. M., & Gia Nhu, N. (Eds.). (2021). Proceedings of International Conference on Machine Intelligence and Data Science Applications. Algorithms for Intelligent Systems. doi:10.1007/978-981-33-4087-9
13. Qiu, Z., Zhu, X., Shi, D., & Kuang, Y. (2020). Recognition of Transmission Line Related Bird Species Based on Image Feature Extraction and Support Vector Machine. 2020 IEEE International Conference on High Voltage Engineering and Application (ICHVE). doi:10.1109/ichve49031.2020.9279508
14. Rajan, R., & A, N. (2021). Multi-label Bird Species Classification Using Transfer Learning. 2021 International Conference on Communication, Control and Information Sciences (ICCISc). doi:10.1109/iccisc52257.2021.9484858
15. Salamon, J., Bello, J. P., Farnsworth, A., & Kelling, S. (2017). Fusing shallow and deep learning for bioacoustic bird species classification. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). doi:10.1109/icassp.2017.7952134

16. Thakur, A., Abrol, V., Sharma, P., & Rajan, P. (2018). Compressed Convex Spectral Embedding for Bird Species Classification. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). doi:10.1109/icassp.2018.8461814
17. Towhid, M. S., & Rahman, M. M. (2017). Spectrogram segmentation for bird species classification based on temporal continuity. 2017 20th International Conference of Computer and Information Technology (ICCIT). doi:10.1109/iccitechn.2017.8281775
18. Xu, W., Zhang, X., Yao, L., Xue, W., & Wei, B. (2020). A Multi-view CNN-based Acoustic Classification System for Automatic Animal Species Identification. Ad Hoc Networks, 102115. doi:10.1016/j.adhoc.2020.102115
19. Yoshihashi, R., Kawakami, R., Iida, M., & Naemura, T. (2017). Bird detection and species classification with time-lapse images around a wind farm: Dataset construction and evaluation. Wind Energy, 20(12), 1983–1995. doi:10.1002/we.2135
20. Zhao, Z., Zhang, S., Xu, Z., Bellisario, K., Dai, N., Omrani, H., & Pijanowski, B. C. (2017). Automated bird acoustic event detection and robust species classification. Ecological Informatics, 39, 99–108. doi:10.1016/j.ecoinf.2017.04.003
21. L. -L. Zhang, Y. Jiang, Y. -P. Sun, Y. Zhang and Z. Wang, "Improvements Based on ShuffleNetV2 Model for Bird Identification," in IEEE Access, vol. 11, pp. 101823-101832, 2023, doi: 10.1109/ACCESS.2023.3314676.
22. Geetanjali S. Mahamunkar, Arvind W. Kiwelekar, and Laxman D. Netak,Deep Learning Model for Black Spot Classification,International Journal of Performability Engineering,18,2020, doi = {10.23940/ijpe.22.03.p8.222230}

23. Shobhanam Krishna and Sumati Sidharth},HR Analytics: Employee Attrition Analysis using Random Forest,year = {2022},journal = {International Journal of Performability Engineering,doi = {10.23940/ijpe.22.04.p5.275281}

24. Rajan Prasad Tripathi, Sunil Kumar Khatri, and Darelle Van Greunen},Relative Examination of Breast Malignant Growth Analysis Utilizing Different Machine Learning Algorithms,year = {2022}, International Journal of Performability Engineering},doi = {10.23940/ijpe.22.06.p4.417425}

25. Poonam Narang, Ajay Vikram Singh, and Himanshu Monga},Hybrid Metaheuristic Approach for Detection of Fake News on Social Media},International Journal of Performability Engineering},doi = {10.23940/ijpe.22.06.p6.434-443}

26. Yerriswamy T and Gururaj Murtugudde},Signature-based Traffic Classification for DDoS Attack Detection and Analysis of Mitigation for DDoS Attacks using Programmable Commodity Switche, International Journal of Performability Engineering},doi = {10.23940/ijpe.22.07.p8.529536}